Histogram (AKA Frequency Chart)

The histogram is a bar graph that summarizes a large amount of data from a process that has been collected over a period.

Use a histogram to....

Display the distribution of measured data in an easy-to-read format, rather than in rows and columns in a table.

Show the centering, variation, and shape of the data.

Show if the data is grouped at points to the right or left of the target value, or the average value, if that is what is required.

How to design a histogram...

1. Data collected should be variable data; that is, it is measured on a continuous scale. Examples: time, dimensions, scores, weights.

2. Gather the data. If you intend to look for patterns in the distribution of the data, or wish to assess the spread (variation) of the data, you should collect at least 50 to 100 data points.

3. Prepare a frequency table from the data.

Example. Here is a listing of test scores:

33	66	70	75	79	82	88	89	100
53	66	71	75	79	84	89	90	
64	67	72	76	81	84	89	90	
65	68	72	77	81	86	89	90	
65	70	73	78	82	88	89	91	

4. Count the number of data points, n, in the set to get the sample size. The number of data points (test scores) is 41: n = 41.

5. Determine the range (highest point minus the lowest point), R, for the sample. The range of test scores is 67 (100-33=67).

6. Determine the number of class intervals, k, needed. If there are less than 50 data points, you should divide the sample into 5-7 classes. If there are between 50 and 100 points, there should be 6-10 classes. If there are between 100 and 250 data points, there should be 7-12 classes, and if there are over 250 points, there should be 10-20 categories (classes). For this example, with 41 data points, there would be 5-7 categories of data.

> Note: Too many classes of data will produce a spread out, flat pattern, and not enough will produce a high and tight pattern.

7. Determine the class width, or the size of each category, by dividing the range, R, by the number of classes, k: Category Width: H = R(67) ÷ k(6) = 11.2 rounded down to 11.

8. To make the first class boundaries, the upper and lower points of each class, add 10 to the lowest data point. This is the first class. Consecutively add the class width, H, to the lowest class boundary until all the range of data points are obtained. Since the upper and lower boundaries are counted, this will make each class contain 11 points, except the last class, which will contain 13 points. Note: each class interval must be mutually exclusive, where every data point will fit into only one class interval.

9. Draw a frequency table, showing the number of data points in each class.

Class	**Class Boundaries**	**Frequency**
1	33-43	1
2	44-54	1
3	55-65	3
4	66-76	13
5	77-87	11
6	88-100	12

10. Draw a histogram from the frequency table. On the vertical (y-axis), draw the frequency scale. Make it long enough to cover the highest frequency count. On the horizontal (x-axis), draw the scale that shows the measured data. For each class of data, draw a bar with the height equal to the frequency for that class.